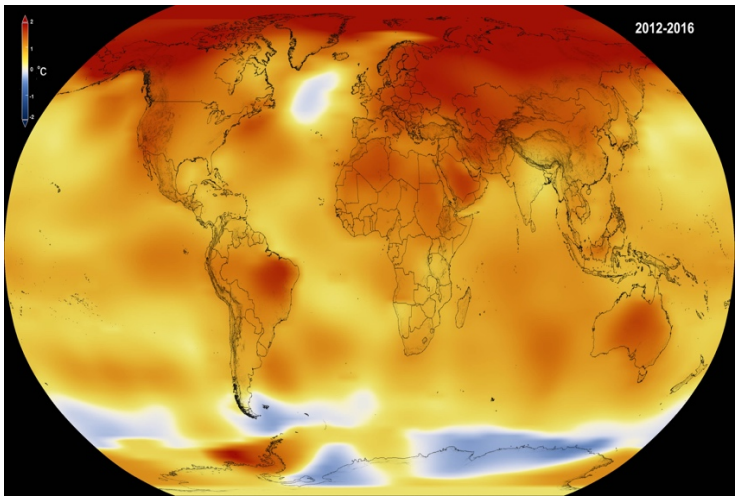
Udacity-Data Analyst Nanodegree – Project 1

EXPLORE WEATHER TRENDS

Tam Nguyen
August 22, 2020



1. Extract Data from Database

1.1. Extract data of the city where I live

First, I live in Houston, United States so first I start to find whether my city is in the database. In order to do that, I wrote an SQL query to retrieve Houston city in the United States.

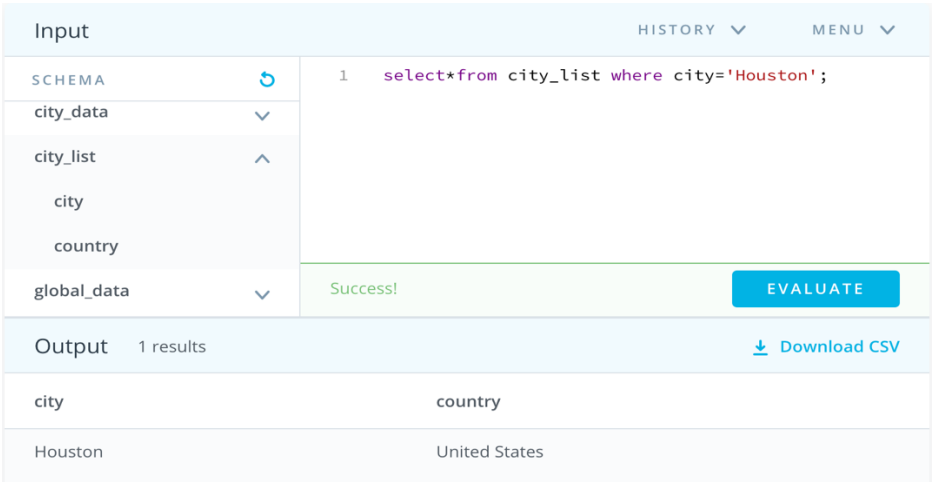


Figure 1.1: SQL Query (Local city)

Next step is extracting all the temperature data including year period for Houston city using SQL query as follows:

Figure 1.2: SQL Query (Houston Data)

This query will return 194 results in output from the year 1982 to 2013.

1.2. Extract global data

I use another SQL query to extract the temperature data for global as follows:

Figure 1.3: SQL Query (Global Data)

There are 266 results in output from the year 1750 to 2015 in the data of global.

2. Data selection and manipulation

2.1.Data selection

The data has been extracted using SQL query and downloaded as CSV files. Then I imported the CSV into Excel for further analyst.

Comparing values from the temperature data of Houston city and global, I noticed that we have more rows in the global dataset which are 266 results compared to Houston city that are 194 results. This is because the global data frame started recording data from 1750 to 2015 while Houston data frame started at 1820 to 2013.

In order to provide the accuracy dataset for comparison, I proceed to choose the common range of years for both which is from 1820 to 2013 to make analytics.

2.2. Data manipulation

To smooth out the results during data manipulation, moving average is the best way to evaluate the data when it compares to yearly average.

Therefore, in the same excel spreadsheet I prepare a column named M-average to each data frame and calculate average temperature on 10 years basic. This is done by calculating the average temperature for the first 10 years (1820-1829). The same process is then repeated from years 1821-1830 and so on.

2.3. Data Visualization

I used Excel to plot the line chart to show the comparison between Houston city average temperature and the global average temperature. There are 3 columns for plotting which are moving average temperature of Houston city, moving average temperature of the global, and the year range. The moving average temperature is on the y-axis while the year range is on the x-axis.

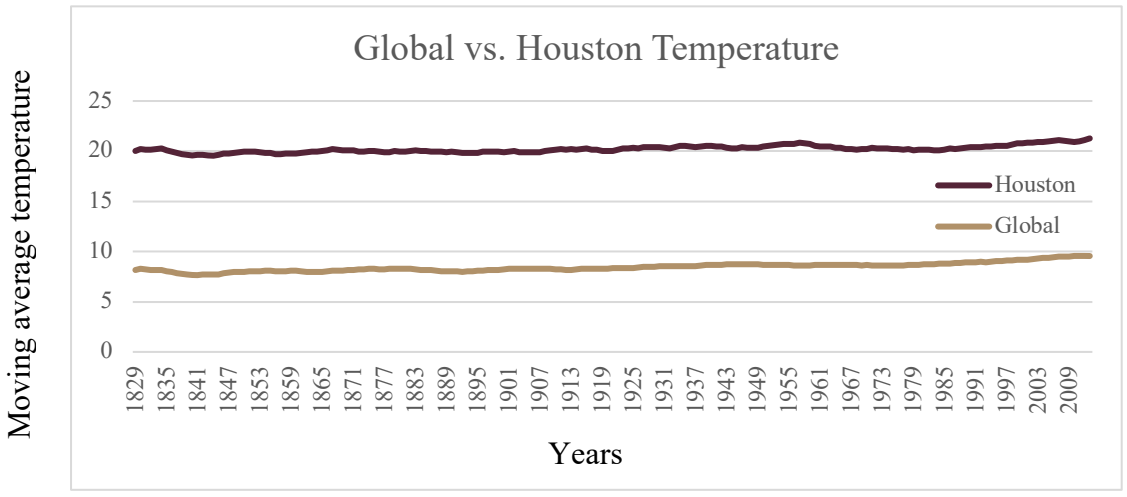


Figure 2.1: Global Vs. Houston Average Temperature

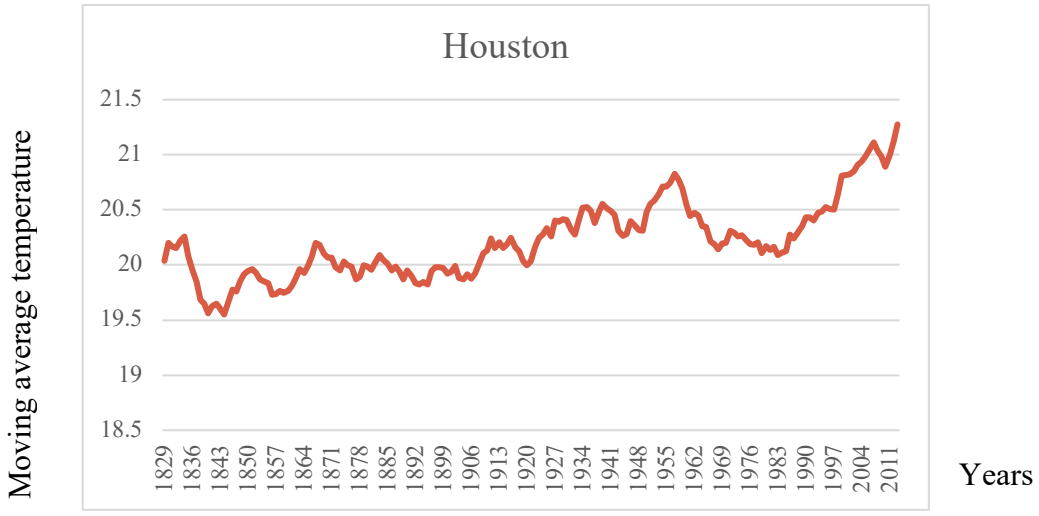


Figure 2.2: Houston Average Temperature

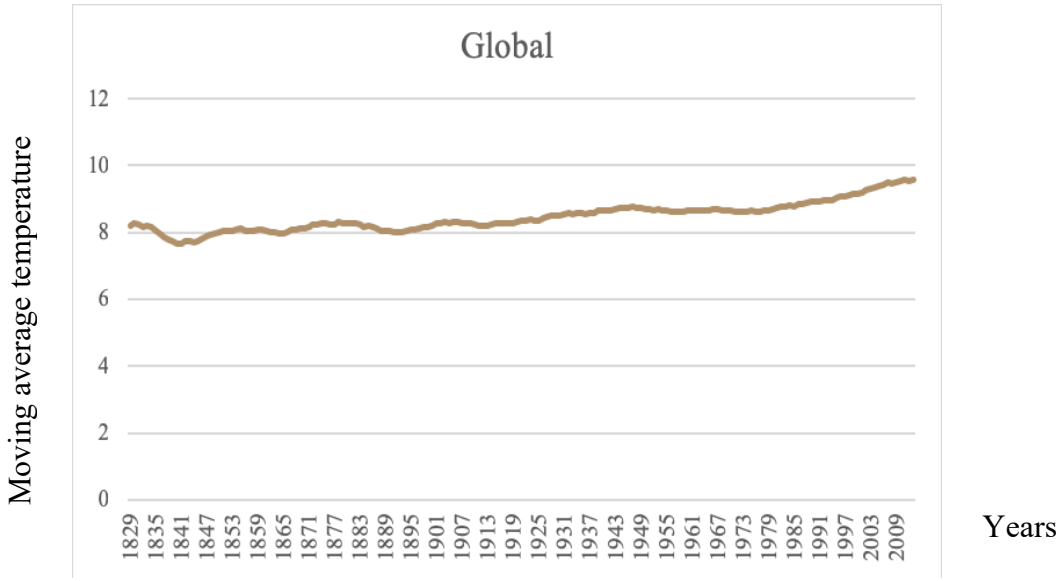


Figure 2.3: Global Average Temperature

3. Observation

Based on the line chart, I observe some similarities and differences in the weather trend between Houston city and the global as follows:

- By noticing the chart, it shows that the temperature is gradually increasing throughout the entire year period in both cases. It means that the earth is getting hotter.
- It is interesting to see that there have been times when there are drops in both global and Houston average temperature. For instance, there is drop for both lines from the range 1841 to 1847.
- The global average temperature shows more consistency in trend line than Houston's average temperature.
- Houston's weather has fluctuated temperature until the year 1990 and then the temperature increases significantly over time and upward.
- Houston's weather is indeed hotter on average compare to the global. This difference has been consisted over time.